

## MEETING REPORT

(2000) are consistent with a similar role for AtRab1b in plant cells. Future experiments will be needed to establish whether the cessation of Golgi movement is specific to the inhibition of Rab1 GTPase function or whether it is a common occurrence whenever ER-Golgi membrane traffic is inhibited. It will be important to establish whether the inhibition of Golgi movement arises from the inhibition of membrane traffic (consistent with the filling-station model) or whether the inhibition of movement is responsible for the inhibition of membrane traffic (consistent with the vacuum cleaner model). In this regard, it will be interesting to investigate the effect on membrane traffic of drugs and mutants that inhibit Golgi movement and to determine the effect on Golgi movement of mutants in other regulatory factors such as Arf1 and Sar1 (which function in the formation of COPI- and COPII-coated vesicles, respectively). It also will be important to establish the site of COPII vesicle formation as Pimpl et al. (2000) have done for COPI. In addition to the first direct evidence for Rab1b GTPase function in ER-Golgi transport in plant cells, Batoko et al. (2000) present an excellent system for uncovering further mechanistic details of trafficking in the secretory pathway.

**Nancy A. Eckardt**  
News and Reviews Editor

## REFERENCES

- Andreeva, A.V., Kutuzov, M.A., Evans, D.E., and Hawes, C.R.** (1998). The structure and function of the Golgi apparatus: a hundred years of questions. *J. Exp. Bot.* **49**, 1281–1291.
- Bannykh, S.I., and Balch, W.E.** (1998). Selective transport of cargo between the endoplasmic reticulum and Golgi compartments. *Histochem. Cell Biol.* **109**, 463–475.
- Batoko, H., Zheng, H.-Q., Hawes, C., and Moore, I.** (2000). A Rab1 GTPase is required for transport between the endoplasmic reticulum and Golgi apparatus and for normal Golgi movement in plants. *Plant Cell* **12**, 2201–2218.
- Bischoff, F., Molendijk, A., Rajendrakumar, C.S.V., and Palme, K.** (1999). GTP binding proteins in plants. *Cell Mol. Life Sci.* **55**, 233–256.
- Boevink, P., Oparka, K., Santa Cruz, S., Martin, B., Betteridge, A., and Hawes, C.** (1998). Stacks on tracks: The plant Golgi apparatus traffics on an actin/ER network. *Plant J.* **15**, 441–447.
- Boevink, P., Martin, B., Oparka, K., Cruz, S.S., and Hawes, C.** (1999). Transport of virally expressed green fluorescent protein through the secretory pathway in tobacco leaves is inhibited by cold shock and brefeldin A. *Planta* **208**, 392–400.
- Dascher, C., and Balch, W.E.** (1994). Dominant inhibitory mutants of ARF1 block endoplasmic reticulum to Golgi transport and trigger disassembly of the Golgi apparatus. *J. Biol. Chem.* **269**, 1437–1448.
- Ellgaard, L., Molinari, M., and Helenius, A.** (1999). Setting the standards: Quality control in the secretory pathway. *Science* **286**, 1882–1888.
- Nebenführ, A., Gallagher, L.A., Dunahay, T.G., Frohlick, J.A., Mazurkiewicz, A.M., Meehl, J.B., and Staehelin, L.A.** (1999). Stop-and-go movements of plant Golgi stacks are mediated by the actin-myosin system. *Plant Physiol.* **121**, 1127–1141.
- Nickel, W., and Wieland, F.T.** (1998). Biosynthetic protein transport through the early secretory pathway. *Histochem. Cell Biol.* **109**, 477–486.
- Palmer, D.J., Helms, J.B., Beckers, C.J., Orci, L., and Rothman, J.E.** (1993). Binding of coatomer to Golgi membranes requires ADP-ribosylation factor. *J. Biol. Chem.* **268**, 12083–12089.
- Pimpl, P., Movafeghi, A., Coughlan, S., Denecke, J., Hillmer, S., and Robinson, D.G.** (2000). In situ localization and in vitro induction of plant COPI-coated vesicles. *Plant Cell* **12**, 2219–2236.
- Scales, S.J., Pepperkok, R., and Kreis, T.E.** (1997). Visualization of ER-to-Golgi transport in living cells reveals a sequential mode of action for COPII and COPI. *Cell* **90**, 1137–1148.
- Schekman, R., and Orci, L.** (1996). Coat proteins and vesicle budding. *Science* **271**, 1526–1533.
- Vitale, A., and Denecke, J.** (1999). The endoplasmic reticulum—gateway of the secretory pathway. *Plant Cell* **11**, 615–628.
- Wee, E.G.-T., Sherrier, D.J., Prime, T.A., and Dupree, P.** (1998). Targeting of active sialyltransferase to the plant Golgi apparatus. *Plant Cell* **10**, 1759–1768.

## Sequencing the Rice Genome

Rice (*Oryza sativa*) is one of the most important crops in the world. Rice, wheat, and maize together account for about half of the world's food production, and rice itself is the principal food of half of the world's population (Sasaki

and Burr, 2000). Rice is the obvious choice for the first whole genome sequencing of a cereal crop. The rice genome is well mapped and well characterized, and it is the smallest of the major cereal crop genomes at an esti-

mated 400 to 430 Mb. The next largest genome of an important cereal crop is that of sorghum, at 750 to 770 Mb, and the wheat genome is ~37 times the size of the rice genome at close to 16,000 Mb (Arumuganathan and Earle,

## MEETING REPORT

1991). Grass genomes, including those of rice, wheat, maize, barley, rye, and sorghum, share a large degree of synteny, making rice an excellent model cereal (Gale and Devos, 1998). Rice is also the easiest of the cereal plants to transform genetically. A genome size of 430 Mb nonetheless represents a daunting task for whole genome sequencing. The rice genome is 3.5 times the size of the *Arabidopsis* genome and the third largest public genome project undertaken to date, behind the human and mouse genomes.

The International Rice Genome Sequencing Project (IRGSP) began in September 1997, at a workshop held in conjunction with the International Symposium on Plant Molecular Biology in Singapore. Scientists from many nations attended the workshop and agreed to an international collaboration to sequence the rice genome. As a result, representatives from Japan, Korea, China, the United Kingdom, and the United States met six months later in Tsukuba to establish the guidelines. The participants agreed to share mate-

rials and to the timely release of physical maps and annotated DNA sequence to public databases. The IRGSP has evolved to include 11 nations, and the IRGSP Working Group, composed of a representative from each participating nation, formulates IRGSP policies and finishing standards. The recent interim IRGSP meeting at Clemson University (September 19 and 20, 2000) in South Carolina was the largest rice genome meeting to date and was attended by more than 70 scientists and administrators from Japan, Taiwan, Thailand, Korea, China, India, Brazil, France, Canada, and the United States. The meeting was organized by Rod Wing, U.S. IRGSP Representative (Clemson University), and chaired by Ben Burr, IRGSP Coordinator (Brookhaven National Laboratory, New York), and Takuji Sasaki, Program Director of the Rice Genome Research Program (RGP) in Japan. Major players in the project include the RGP; the CCW, a collaboration between the Clemson University Genomics Institute (CUGI), Cold Spring Harbor Laboratory, and the Washing-

ton University Genome Sequencing Center; the Institute for Genome Research (TIGR) in Rockville, MD; and the Plant Genome Initiative at Rutgers University (PGIR). Various additions and/or changes in IRGSP members were noted at the meeting. Brazil became the newest member and was represented by Antonio Costa de Oliveira of the Universidad Federal de Pelotas, who proposed to work on chromosome 12. Canada representative Thomas Bureau of McGill University proposed switching from work on chromosome 2 to coordinating activities on chromosome 9 with Thailand. India, previously an unfunded member of the IRGSP, has a new Rice Genome Program (represented by Akhilesh Tyagi of the University of Delhi and Nagendra Singh of the Indian Agricultural Research Institute) and will begin work on chromosome 11. A full list of participating countries and institutions, including URLs of sites offering information relevant to the IRGSP, is provided in Table 1.

Rice genome sequencing is being conducted along the same lines as

**Table 1.** Rice Sequencing Participants and Chromosome Assignments

Site	Chromosome	URL <sup>a</sup>
Rice Genome Research Program (RGP; Japan)	1, 6, 7, 8	<a href="http://rgp.dna.affrc.go.jp/index.html">http://rgp.dna.affrc.go.jp/index.html</a>
Korea Rice Genome Research Program (Korea)	1	<a href="http://bioserver.myongji.ac.kr/ricemac.html">http://bioserver.myongji.ac.kr/ricemac.html</a>
CCW (United States)	3, 10	
CUGI (Clemson University)		<a href="http://www.genome.clemson.edu/">http://www.genome.clemson.edu/</a>
Cold Spring Harbor Laboratory		<a href="http://www.cshl.org/">http://www.cshl.org/</a>
Washington University Genome Sequencing Center		
TIGR (United States)	3, 10	<a href="http://www.tigr.org/tdb/rice/">http://www.tigr.org/tdb/rice/</a>
PGIR (United States)	10	<a href="http://pgir.rutgers.edu/">http://pgir.rutgers.edu/</a>
University of Wisconsin (United States)	11	
National Center for Gene Research	4	<a href="http://www.ncgr.ac.cn/index.html">http://www.ncgr.ac.cn/index.html</a>
Chinese Academy of Sciences (China)		
Indian Rice Genome Program (University of Delhi)	11	
Academia Sinica Plant Genome Center (Taiwan)	5	<a href="http://genome.sinica.edu.tw/index_e.htm">http://genome.sinica.edu.tw/index_e.htm</a>
Genoscope (France)	12	<a href="http://www.genoscope.cns.fr/">http://www.genoscope.cns.fr/</a>
Universidade Federal de Pelotas (Brazil)	12	
Kasetsart University (Thailand)	9	
McGill University (Canada)	9	
John Innes Centre (United Kingdom)	2	

<sup>a</sup> URLs are listed only for sites that currently provide information relevant to rice genome sequencing.

## MEETING REPORT

numerous other large-scale genome sequencing projects. Large insert genomic libraries, used as the primary sequencing templates, are constructed in bacterial artificial chromosomes (BACs) or P1-derived artificial chromosomes (PACs). Sequencing of the rice genome is being performed mainly from genomic BAC or PAC libraries created from the Nipponbare variety, which was chosen as the common template throughout the IRGSP; China, working on the sequencing of chromosome 4, is the only IRGSP member to use a different variety, *indica Guang Lu Ai 4* (Sasaki and Burr, 2000). Budiman (1999), in a report accessible through the CUGI website, presents a complete description of the preparation of two deep-coverage rice BAC libraries (25-fold genome coverage) used by the IRGSP.

## PROGRESS REPORTS

The meeting began with progress reports from several IRGSP members. Takuji Sasaki presented the progress of the RGP (Japan) on the short arm of chromosome 1. To date, 98 PACs/BACs corresponding to ~12 Mb of sequence have been completed and released to the DNA Data Bank of Japan. At least 53 of the PACs/BACs have been annotated. RGP is also working on an expressed sequence tag (EST) map and has generated ~5000 EST markers, which will be added to the rice database (Integrated Rice Genome Explorer, or INE; Sakata et al., 2000); these data are available at the RGP website (see also Yamamoto and Sasaki, 1997). Robin Buell (TIGR) and Rod Wing (CUGI) reported on progress in sequencing chromosomes 10 and 3. CUGI has released 36 Mb of rice sequence tagged connectors (STCs), and the CCW has released 6 Mb of sequence to GenBank from chromosomes 3 and 10 and recently published

a report on rice sequence data mining (Mao et al., 2000). STCs are sequences from BAC ends that are sequenced at random and used to create minimal tiling paths for complete sequencing (Venter et al., 1996; Mahairas et al., 1999). The TIGR rice website has been updated to include nearly 7 Mb (of an estimated 9.6 to 11.5 Mb) of sequence in progress from the bottom arm of chromosome 10. TIGR has just begun sequencing chromosome 3 and has at least 4.8 Mb of an estimated 28 Mb currently in production. It should be noted that the U.S. groups (CCW, TIGR, and PGIR) together have completed sequencing ~3.2 Mb (with annotation completed on a large portion of this); the remainder is in various stages of Phase I and II unfinished sequence. TIGR has created a new database and a new display on its website, which includes the ability to search for gene names, and preliminary annotation of unfinished BACs. This group is working on a rice gene index that will be linked to other plant gene indices on the website and to an orthologous gene alignment database for identification of putative orthologs and paralogs, with graphic displays of alignments between species. A current problem is the lack of functional genomics data: many sequences are homologous to sequences in Arabidopsis that are hypothetical genes, with no information on expression or possible function. Updates on progress in sequencing of all the various IRGSP members, as well as links to other IRGSP websites, can be found at the RGP website (<http://rgp.dna.affrc.go.jp/Seqcollab.html>).

#### PHYSICAL MAPPING AND THE CONSTRUCTION OF "SEQUENCE-READY" MINIMAL TILING PATHS

Fingerprinting and physical mapping are used to create minimal tiling paths and to anchor BAC clones to physical

positions along the length of a chromosome. A sequence-ready BAC contig is a contiguous set of minimally overlapping BAC clones that has been anchored to a position along the length of a particular chromosome (described in detail in Zhang and Wing, 1997). Presentations during the opening session on physical mapping and a physical mapping workshop provided a good overview of the various techniques and procedures used during this process. Gernot Presting (formerly of CUGI and now at the Novartis Agricultural Discovery Institute in San Diego) described the CUGI physical mapping project (now led by Eric Fang and others at CUGI). The project involves fingerprinting of HindIII and EcoRI BAC libraries, assembling the fingerprinted BACs into contigs, anchoring of the BACs onto the physical map with DNA gel restriction fragment length polymorphism (RFLP) and BAC end sequence analysis, and connecting and extending of contigs by chromosome walking. Another project involves the mapping of plant ESTs onto the rice physical map. Information on the ESTs is being integrated into the rice physical map and made accessible on the CUGI website.

Assembly of the fingerprinted BACs into contigs is greatly aided by the use of software called FingerPrinted Contigs (FPC) developed by Soderlund et al. (2000). Soderlund described how she and others at CUGI are working on developing ways of using FPC in conjunction with data from markers (DNA gel blot analysis), fluorescence in situ hybridization (FISH) and optical mapping (when available), and the BAC end sequence (STC) database to improve the efficiency and reliability of creating sequence-ready BAC contigs. Eric Fang of CUGI described techniques for extending BAC contigs and closing gaps as part of the process to create a sequence-ready framework for the entire genome. One method described was the use of overgo probes. In this procedure, a pair of 24-bp sequences

## MEETING REPORT

that contain an 8-bp overlap is designed from BAC end sequences. The 24-bp sequences are joined to create a 40-bp "overgo," which is used to probe a high-density BAC library filter to find additional BAC clones that may extend a contig (software available at <http://genome.wustl.edu/gsc/overgo/overgo.html>).

FISH and optical mapping are two other methods that may be used to aid genome sequencing projects. Jiming Jiang (Department of Horticulture, University of Wisconsin) presented data to illustrate the application of FISH to rice physical mapping. The length of the pachytene chromosome structure in rice makes it particularly amenable to FISH, and resolution and sensitivity are comparable to results that have been obtained from human chromosomes. Jiang showed how FISH could be used to easily identify rice chromosomes, to determine the chromosome location of uncertain clones, and to determine the physical nature of large linkage gaps, which could facilitate sequence closing at chromosome ends and telomere regions. Jiang's group previously reported that this technique would be valuable for characterizing BAC clones that contain complex repetitive DNA sequences such as those found in rice (Jackson et al., 1999). Sally Leong (United States Department of Agriculture Agricultural Research Service Plant Disease Resistance, Madison, WI) described the optical mapping technique developed by David Schwartz and colleagues at the University of Wisconsin. This technique uses fluid flow capillary action to extend and align DNA molecules onto a specially prepared glass surface. DNA that is extended and fixed to the surface is then digested with restriction enzymes, and fluorescence microscopy imaging is used to map an ordered array of fragments. Schwartz's group has used optical mapping to create whole genome restriction maps of the microorganisms *Deinococcus radiodurans* (Lin et al., 1999) and *Plasmodium falciparum* (Lai et

al., 1999), and they received a National Science Foundation Plant Genome Award in 1999 to support the construction of an optical restriction map of the rice genome. FISH and optical mapping could significantly enhance physical mapping and sequencing of difficult regions, such as centromeres and regions containing highly repetitive sequence, although their expense currently limits the extent to which they are being used in genome sequencing.

### SEQUENCING, FINISHING, AND ANNOTATION

The Rice Genome Research Program (Japan) uses a shotgun approach to sequence PAC or BAC clones. With this procedure, individual PAC/BAC clones (100 to 200 kb) from a sequence-ready contig are shattered by sonication or nebulization, and the fragments are subcloned to produce a shotgun library with an average insert size of 1 to 3 kb. Clones from the shotgun library are then sequenced at random to provide the desired degree of "coverage" of the total sequence. For example, to provide for fourfold coverage of a 120-kb BAC, at least 1200 clones from a shotgun library would be sequenced at random (assuming 400 bp per sequence read).

After sequencing, software such as PHRED and PHRAP is used to order the subclone sequences and reassemble the entire BAC sequence. PHRED, developed by Phil Green and Brent Ewing at the University of Washington, reads DNA sequencer trace data, calls bases, and assigns quality values to the bases; PHRAP is a program developed by Phil Green for assembling shotgun DNA sequence data. Of course, it is rare that an entire BAC will be assembled without gaps from shotgun sequences. Doug Johnson (Washington University Genome Sequencing Center), Melissa de la Bastide (Cold Spring

Harbor Laboratory), Robin Buell (TIGR), and Apichart Vanavichit (Kasetsart University, Thailand) presented information on the finishing process and discussed ways to deal with problem regions and filling gaps. Problem regions in sequencing include highly repetitive sequence and AT-rich and GC-rich regions. The rice genome carries a large amount of repetitive sequence. Problems in these areas often can be overcome by switching the sequencing chemistry; de la Bastide presented a list of various sequencing kits and reagents that have been used successfully at Cold Spring Harbor Laboratory to sequence through difficult areas. She discussed two other ways of dealing with particularly recalcitrant regions: the use of small insert libraries and transposons. Small insert libraries are created by physical shearing and subcloning of a template that spans the region of difficulty. Transposons also may be used to break up and thereby aid the sequencing of a difficult region, which is achieved by random insertion of transposons into a difficult clone. De la Bastide and Steven Salzberg of TIGR also presented a list of finishing resources available on the Internet; these included the Washington University Finisher Related Technology development (<http://genome.wustl.edu/gsc/TechD/finishing.htm>), the University of Washington finishing protocol (<http://www.genome.washington.edu/UWGC/finish.htm>), and the Sanger Center finishing software ([www.sanger.ac.uk/Software/](http://www.sanger.ac.uk/Software/)).

Sequence annotation was discussed by Steven Salzberg (TIGR) and Todd Wood (formerly of CUGI and now at Bryan College in Dayton, TN). Salzberg presented information on the eukaryotic gene finder program GLIMMER, which he and others have been "training" to work on rice, and a system called MUMmer, which is used to find repetitive sequences. This software is available on the TIGR website (<http://www.tigr.org/softlab>). Wood spoke about the use of GenTerpret annotation software by Rabbithutch Biotechnology

## MEETING REPORT

Corp. (<http://www.rabbithutch.com>). Using this software, researchers annotating four clones at the top of the short arm of rice chromosome 10 located 89 protein-coding genes. Approximately 50% were hypothetical (25% hypothetical with an EST), and 52% had a homolog in Arabidopsis. Wood is poised to help with phase II annotation, which will include the integration of functional genomics data, such as information on tissue distribution, subcellular localization, developmental stages, physiology, and post-transcriptional and post-translational regulation.

#### RELEASE OF MONSANTO RICE SEQUENCE DATA AND INTEGRATION WITH IRGSP

Monsanto announced on April 4, 2000, that the company had completed a working draft of the rice genome, which would be made available to the IRGSP ([http://www.monsanto.com/monsanto/mediacenter/2000/00apr4\\_rice.html](http://www.monsanto.com/monsanto/mediacenter/2000/00apr4_rice.html)). The company has proposed an agreement whereby data and materials could be transferred to IRGSP members. Gerard Barry of Monsanto stated that the company wished to facilitate the release of its rice sequence data and materials to the public but also to avoid the creation of rival sequencing efforts that could undermine both public and private efforts. The report from the working group meeting suggested that all sides were close to a workable agreement, but it remains for many of the individual institutions involved to sign a contract. Many scientists and administrators at public institutions in the United States should watch the unfolding of this agreement with interest. In recent years, many universities and public institutions have attempted to foster collaborations and ties with industry yet have often balked at agreements that private companies see as reasonable and necessary to protect their business interests.

Tomoya Baba (RGP) and Brad Barbazuk (Monsanto) gave presentations regarding the integration of RGP and CUGI data, respectively, with the Monsanto rice physical map. Barbazuk gave an overview of Monsanto's rice genomics resources. Monsanto has 50,895 BACs with HindIII fingerprints and more than 135,000 STCs. Physical mapping was conducted using in silico physical map data available from the RGP (i.e., BlastN was used to match BAC end sequences to RFLP markers and create contigs; no wet chemistry physical mapping was performed at Monsanto). Approximately 3400 BACs were selected, forming contigs that represent 393 Mb of the rice genome. Monsanto has transferred these ~3400 BAC clones to the RGP, along with sequence information for more than 3300 clones, in silico physical map data for almost 2000 clones, and more than 125,000 STCs. Barbazuk also discussed efforts to merge the Monsanto physical map with CUGI data, such as associating CUGI and Monsanto contigs by identifying CUGI STCs with high confidence hits to Monsanto BAC clones. Baba presented an analysis of the Monsanto data and initial attempts at RGP to integrate the data into the RGP physical map. It was clear that integration would be a difficult task and that there were some problems with the Monsanto data (perhaps stemming from the physical map being derived solely in silico). Nonetheless, it was widely agreed that the Monsanto data ultimately would be an enormous help to the rice sequencing project and would allow for much more timely completion of the rice genome.

One problem with the Monsanto agreement, which was discussed at length in several sessions, is the current stipulation that all sequence derived from Monsanto data not be published until it is combined with IRGSP sequence in completed BACs or PACs. Some members of the IRGSP currently

submit unfinished sequence assemblies greater than 2 kb to the HTGS (for high throughput genome sequence) division of GenBank (see <http://www.ncbi.nlm.nih.gov/HTGS/>). A tentative compromise was discussed that would permit the release of Monsanto data when combined with IRGSP-generated sequence in phase II HTGS submissions.

#### GENOSCOPE PROPOSAL: WHOLE GENOME SHOTGUN SEQUENCING OF THE RICE GENOME

Francis Quetier (France) presented a Genoscope proposal for global shotgun sequencing of the rice genome. The proposal called for stopping the current BAC-by-BAC approach and performing whole genome shotgun sequencing with the goal of obtaining an additional four-fold coverage according to the Bermuda sequencing standards (the international human genome sequencing community held meetings in Bermuda in 1996 and 1997 to set standards for DNA sequence; described at <http://www.gene.ucl.ac.uk/hugo/bermuda2.htm>). This would result in an estimated 1.75 million clones and 3.5 million BAC ends to sequence, based on 430 total Mb and 600-bp sequencing reads on both ends of each insert at 85% efficiency. Clones would be distributed among IRGSP sequencing groups on a basis roughly proportional to their current chromosome claim, and each group would submit sequence to a common public database accessible to all groups. Either in parallel with the shotgun sequencing or after it was completed, group members would return to the current BAC-by-BAC approach to complete various chromosomal regions. IRGSP Coordinator Ben Burr stated that there would be ongoing discussion of the proposal among members of the working group and indicated that it may be feasible, and desirable, to integrate

## MEETING REPORT

various aspects of the proposal into IRGSP policy.

### POLICY DISCUSSION AND REVISION OF FINISHING STANDARDS

The goal of the RGP is a complete and accurate sequence of the entire genome. "Complete" was originally defined as less than one error in 10,000 bases, consistent with the Bermuda standards. The measure of completeness was previously considered to be a PHRED score (quality value) of 40 or greater. Sasaki presented empirical evidence showing a quality value of 30 to be consistent with this level of accuracy. There was general agreement on revising the finishing standards to this value to speed the release of sequence while maintaining high-quality data. There was considerable debate regarding a number of other revisions. One of these was whether or not small gaps (such as occur in GC-rich regions) could be left in "completed" sequence representing a single contig. On this point, Sasaki presented evidence from the RGP that these regions are likely to contain open reading frames and that every effort should be made to close gaps before stating that a contig is complete. Others argued that it may be more desirable to release large contigs with small gaps, because closing the gaps in these regions is likely to take a considerable amount of time and effort. Revision of the finishing standards was still under discussion after the meeting. Finally, Ben Burr noted that IRGSP currently is not in line with the Bermuda standards in that groups are only encouraged and not required to release preliminary sequence information. A compromise was discussed that would require submission of phase II data and encourage phase I release. The IRGSP working group is preparing

a modified data release policy that will be finalized and released by early 2001.

Robin Buell (TIGR) presented data on the effect of fourfold rather than eightfold coverage on the quality of HTGS phase II sequence released to the public before closure. Her group compared the quality of sequence that would be obtained from fourfold versus eightfold coverage by using data from three rice BACs that have been completed. There was a substantial difference in the quality of the data. The fourfold coverage yielded 42 contigs, and the largest contig extended 14.8 kb; whereas the eightfold coverage of the same region yielded only 14 contigs, and the largest contig extended nearly 35 kb. For HTGS submissions, fourfold coverage was predicted to miss, on average, 16% of the sequence that would be obtained through eightfold coverage. Thus, release of fourfold instead of eightfold coverage data could be an important factor for end users before closure and makes a convincing argument for the release of combined Monsanto and IRGSP data at phase II (which together represent at least eightfold coverage).

### CONCLUSION

Sequencing of the rice genome is a monumental task. To date, ~3.5% of the genome has been completed (15 of 430 Mb) and another 3 to 5% is in production. Nonetheless, the data that have been released have already provided valuable information on genome structure and organization (see, e.g., Mao et al., 2000), much of which will apply to other cereal crops and to monocots in general. A major part of the nuclear genomes of most plants, and indeed many eukaryotes, is composed of repetitive DNA elements. Repetitive DNA is estimated to constitute at least 50% of the rice genome and as

much as 70% of the maize genome (Nagano et al., 1999). Complete sequencing of the rice genome will provide valuable information on the effect of repetitive elements on genome organization and evolution in plants. The IRGSP also constitutes a proving ground for sequencing and finishing methods for complex genomes, which will provide excellent resources for other eukaryotic genome sequencing projects in the future.

**Nancy A. Eckardt**  
News and Reviews Editor

### REFERENCES

- Arumuganathan, K., and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **3**, 208–218.
- Budiman, M.A. (1999). Construction and characterization of a BAC library of *Oryza sativa* L. ssp. *japonica* cv. Nipponbare for genomic studies. URL <http://www.genome.clemson.edu/where/budiman/index.html>.
- Gale, M.D., and Devos, K.M. (1998). Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* **95**, 1971–1974.
- Jackson, S.A., Dong, F., and Jiang, J. (1999). Digital mapping of bacterial artificial chromosomes by fluorescence in situ hybridization. *Plant J.* **17**, 581–587.
- Lai, Z., et al. (1999). A shotgun optical map of the entire *Plasmodium falciparum* genome. *Nat. Genet.* **23**, 309–313.
- Lin, J., Qi, R., Aston, C., Jing, J., Anantharaman, T.S., Mishra, B., White, O., Venter, J.C., and Schwartz, D.C. (1999). Whole genome shotgun optical mapping of *Deinococcus radiodurans*. *Science* **285**, 1558–1562.
- Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and Hood, L. (1999). Sequence-tagged connectors: A se-

quence approach to mapping and scanning the human genome. *Proc. Natl. Acad. Sci. USA* **96**, 9739–9744.

- Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Woo, S.S., Sasinowski, M., Goff, S., Dean, R.A., and Wing, R.A.** (2000). Rice transposable elements: A survey of 73,000 sequence-tagged-connectors (STCs). *Genome Res.* **10**, 982–990.
- Nagano, H., Wu, L., Kawasaki, S., Kishima, Y., and Sano, Y.** (1999). Genomic organization of the 260 kb surrounding the *waxy* locus in a japonica rice. *Genome* **42**, 1121–1126.
- Sakata, K., Antonio, B.A., Mukai, Y., Nagasaki, H., Sakai, Y., Makino, K., and Sasaki, T.** (2000). INE: A rice genome database with an integrated map view. *Nucleic Acids Res.* **28**, 97–101.
- Sasaki, T., and Burr, B.** (2000). International Rice Genome Sequencing Project: The effort to completely sequence the rice genome. *Curr. Opin. Plant Biol.* **3**, 138–141.
- Soderlund, C., Humphray, S., Dunham, A., and French, L.** (2000). Contigs built with fingerprints, markers and FPC V4.7. *Genome Res.*, in press.
- Venter, J.C., Smith, H.O., and Hood, L.** (1996). A new strategy for genome sequencing. *Nature* **381**, 364–366.
- Yamamoto, K., and Sasaki, T.** (1997). Large-scale EST sequencing in rice. *Plant Mol. Biol.* **35**, 135–144.
- Zhang, H.-B., and Wing, R.A.** (1997). Physical mapping of the rice genome with BACs. *Plant Mol. Biol.* **35**, 115–127.